HSE Deep Learning

Image Captiononing

Ashuha Arseniy[1,2]

Bayesian Research Group[1], MIPT[2]
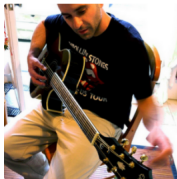
$p(\mathbf{B}|\mathbf{A})$**yesgroup.ru**

MIPT
MOSCOW INSTITUTE
OF PHYSICS AND TECHNOLOGY

ars-ashuha.ru/slides

November 16, 2016

Motivation

- Just for fun
  http://cs.stanford.edu/people/karpathy/deepimagesent/rankingdemo/

"man in black shirt is playing guitar."

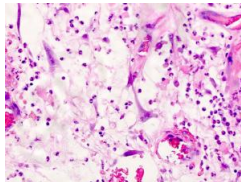"construction worker in orange safety vest is working on road."
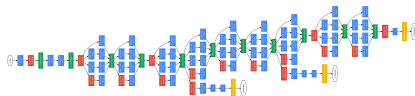
"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."
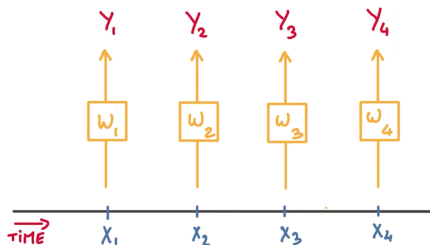
- Medical application

What should we do?

- ▶ We want to make image captioning
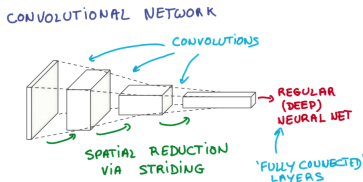- ▶ Make image description

- ▶ Make words description

- ▶ Generative Model

Convolution Neural Nets

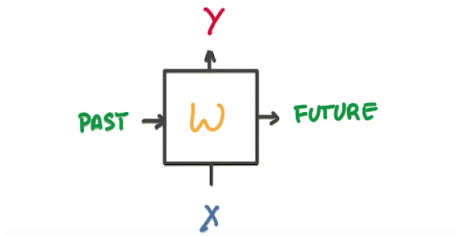- We want to make image captioning
- Make image description

- Make words description
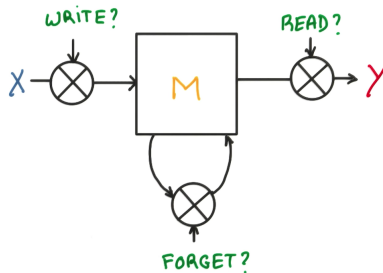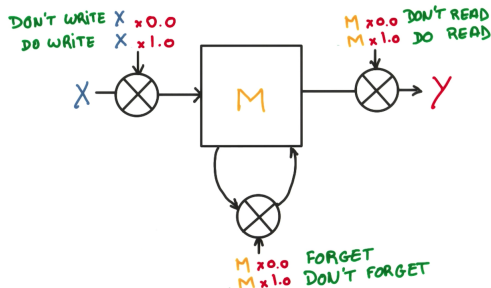

CONVOLUTIONAL NETWORK

CONVOLUTIONS

REGULAR
(DEEP)
NEURAL NET

SPATIAL REDUCTION
VIA STRIDING

'FULLY CONNECTED'
LAYERS

- Generative Model

Recurrent Neural Nets

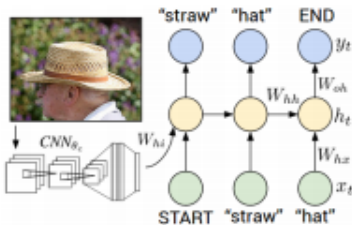- Conception

- LSTM

Recurrent Neural Nets

- LSTM

- Embeding

The domestic cat is   a little tiger ....
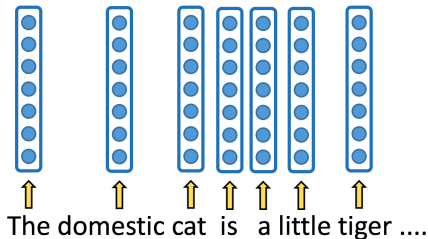
Image Captioning

▶ Captioning
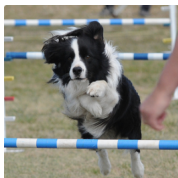
"black and white dog jumps over bar."

▶ Captioning?

http://cs.stanford.edu/people/karpathy/cvpr2015.pdf

Image Captioning
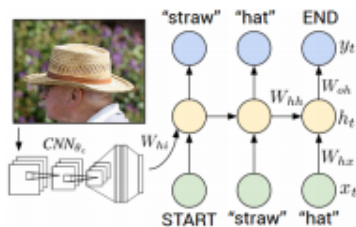
http://mybinder.org/repo/ars-ashuha/caption_binder